

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:54 ; Search time 91.75 Seconds

(without alignments)  
16.543 Million cell updates/sec

Title: US-09-331-631a-8\_COPY\_33\_79

Perfect score: 275

Sequence: 1 GDDDPKRYEDCRRRCCEMDT.....QCESCKSKOYGEKDQQRHR 47

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	588	1	VCLB_GOSHI
2	140	50.9	605	1	VCLA_GOSHI
3	75	27.3	47	1	AGRP_LOFCY
4	67.5	24.5	471	1	VSWA_TRYBR
5	62	22.5	360	1	MJDL_HUMAN
6	61	22.2	189	1	GRPE_HELPY
7	61	22.2	1905	1	TAGB_DICDI
8	59.5	21.6	573	1	GLBL_MAIZE
9	59	21.5	1339	1	DPOA_TRYBR
10	58.5	21.3	443	1	FBL4_CRIGR
11	58.5	21.3	543	1	MIG1_KLUMA
12	58.5	21.3	1085	1	YARA_SCHPO
13	58	21.1	338	1	LEGB_PEA
14	57.5	20.9	443	1	FBL4_HUMAN
15	57.5	20.9	443	1	FBL4_MOUSE
16	57.5	20.9	461	1	KLB2_ECOLI
17	57.5	20.9	650	1	Y096_MYCE
18	57	20.7	1898	1	TRHY_HUMAN
19	56	20.4	301	1	PHLP_RAT
20	56	20.4	481	1	KNIR_DROVI
21	56	20.4	1816	1	LMA4_HUMAN
22	55.5	20.2	2327	1	CCAB_MOUSE
23	55.5	20.2	2336	1	CCAB_RAT
24	55.5	20.2	2339	1	CCAB_HUMAN
25	55.5	20.2	2339	1	CCAB_HUMAN
26	55	20.0	462	1	U2R2_MOUSE
27	55	20.0	540	1	PNR_DROME
28	55	20.0	812	1	PLMN_MOUSE
29	55	20.0	3344	1	POLG_PRSVH
30	54.5	19.8	707	1	DREB_RAT
31	54.5	19.8	841	1	IEG3_MCMVS
32	54	19.6	1744	1	TANA_XENLA
33	54	19.6	2254	1	CCAG_RAT

34	53.5	19.5	405	1	IF4A_CRYPV
35	53.5	19.5	846	1	ITBX_DROME
36	53.5	19.5	1391	1	MSR2_DROXY
37	53	19.3	60	1	PHLP_HUMAN
38	53	19.3	342	1	CYSP_SCHJA
39	53	19.3	400	1	T2PB_YEAST
40	53	19.3	466	1	CYP8_CAEEL
41	53	19.3	691	1	CNGL_CANFA
42	53	19.3	892	1	BN14_YEAST
43	53	19.3	1549	1	TRHY_SHEEP
44	52.5	19.1	394	1	YEL8_YEAST
45	52.5	19.1	538	1	GEL1_YEAST

#### ALIGNMENTS

```

RESULT 1
ID      VCLB_GOSHI      STANDARD:      PRT:      588 AA.
AC      P09801:
DT      01-MAR-1989 (Rel. 10, Created)
DT      01-MAR-1988 (Rel. 10, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS      Gossypium hirsutum (Upland cotton).
OC      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC      Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC      Malvales; Malvaceae; Gossypium.
RN      [1]
RP      'Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RA      "Developmental biochemistry of cottonseed embryogenesis and
RT      germination. XVIII. CDNA and amino acid sequences of the members of
RT      the storage protein families."
RL      Plant Mol. Biol. 7:475-489(1986).
CC      -1- FUNCTION: SEED STORAGE PROTEIN.
CC      -1- SUBCELLULAR LOCATION: CYTOLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC      BODIES.
CC      -1- SIMILARITY: TO OTHER 75 SEED STORAGE PROTEINS (PHASCOLIN, VICILIN,
CC      CONVICILIN, CONGICININ, ETC.).
CC
CC      -----
CC      THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb.ch).
CC      -----
DR      EMBL: M16891; AAA33071.1; -.
DR      PIR: A30838; FWCNAB.
DR      HSSP: P50477; ICAX.
DR      INTERPRO: IPR001113; -.
DR      PFM: PF00546; Seedstore_7s; 1.
KW      Seed storage protein; signal.
FT      SIGNAL
FT      CHAIN
FT      SEQUENCE
SQ      588 AA; 69729 MW; 63B699B29AB8ADB CRC64;

Query Match      100.0%; Score 275; DB 1; Length 588;
Best local similarity 100.0%; Pred. No. 1.4e-22;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GDDDPKRYEDCRRRCCEMDTGRGKEDQOQCESCKSKOYGEKDQQRHR 47
DB      33 GDDDPKRYEDCRRRCCEMDTGRGKEDQOQCESCKSKOYGEKDQQRHR 79

RESULT 2
ID      VCLA_GOSHI      STANDARD:      PRT:      605 AA.

```

```

AC P09799;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chian C.A., Borrito K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR00113; -
DR PFM: PF00546; Seedstore.75; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A
FT SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;
SO
Query Match 50.9%; Score 140; DB 1; Length 605;
Best Local Similarity 54.5%; Pred. No. 4.1e-08;
Matches 24; Conservative 11; Mismatches 7; Indels 2; Gaps 1;
QY 2 DDDPKRYEDCRRRCMDTRGQKEOQCEESCKSOYGEKDOQR 45
DB 34 EDDPQRYEDCKRRCOLRTROTEDQKCDRSETQL--KEEQQR 75
RESULT 3
AGRP_LUFCY STANDARD: PRT: 47 AA.
ID AGRP_LUFCY
AC P56568;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 6.5 KDa ARGININE/GLUTAMATE-RICH POLYPEPTIDE (6.5K-AGRP).
OS Luffa cylindrica (Smooth loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Cucurbitales; Cucurbitaceae; Luffa.
RN [1]
RP SEQUENCE.
RT TISSUE=SEED;
RC MEDLINE=97357433; PubMed=9214759;
RX Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.;
RT "Primary structure of 6.5K-arginine/glutamate-rich polypeptide from
RT the seeds of sponge gourd (Luffa cylindrica)."
RL Biosci. Biotechnol. Biochem. 61:984-988(1997).
CC -1- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON
CC RESERVES DURING GERMINATION AND SEEDLING GROWTH.
CC -1- MASS SPECTROMETRY: MW=5693.39; METHOD=MALDI.

```

```

CC -1- SIMILARITY: SOME: TO 7S SEED STORAGE PROTEINS.
KW Seed storage protein.
FT DISULFID 12 33
FT DISULFID 16 29
SQ SEQUENCE 47 AA; 5698 MW; 588B0EC82273AC05 CRC64;
Query Match 27.3%; Score 75; DB 1; Length 47;
Best Local Similarity 32.4%; Pred. No. 0.035;
Matches 12; Conservative 10; Mismatches 15; Indels 0; Gaps 0;
QY 5 PPKRYEDCRRRCMDTRGQKEOQCEESCKSOYGEKD 41
DB 5 PRTYECRCRVQVAHEGVERQRCQVCEKRLRERE 41
RESULT 4
VSWA_TRYBR STANDARD: PRT: 471 AA.
ID VSWA_TRYBR
AC P20946;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE VARIANT SURFACE GLYCOPROTEIN WRATAT A PRECURSOR (VSG).
OS Trypanosoma brucei rhodesiense.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WRATAT 1;
RX MEDLINE=90290520; PubMed=2357229;
RA Reddy L.V., Hall T., Donelson J.E.;
RT "Sequences of three VSG mRNAs expressed in a mixed population of
RT Trypanosoma brucei rhodesiense."
RL Biochem. Biophys. Res. Commun. 169:730-736(1990).
CC -1- FUNCTION: VSG FORMS A COAT ON THE SURFACE OF THE PARASITE. THE
CC TRYPAOSOME EVADES THE IMMUNE RESPONSE OF THE HOST BY EXPRESSING
CC A SERIES OF ANTIGENICALLY DISTINCT VSGS FROM AN ESTIMATED 1000
CC VSG GENES.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC A SOLUBLE FORM IS RELEASED FROM RUPTURED CELLS BY THE ACTION OF A
CC PI-PLC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M33823; AAA30316.1; -
DR PIR: A35480; A35480.
KW Glycoprotein; Antigen; Trypanosomiasis; GPI-anchor; Membrane; Signal.
FT SIGNAL 1 18
FT CHAIN 19 454 VARIANT SURFACE GLYCOPROTEIN WRATAT A.
FT PROPEP 455 471 HYDROPHOBIC. REMOVED DURING MATURATION.
FT LIPID 454 454 GPI-ANCHOR (BY SIMILARITY).
FT CARBOHYD 61 61 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 471 AA; 50537 MW; 57EC0821042376D CRC64;
Query Match 24.5%; Score 67.5; DB 1; Length 471;
Best Local Similarity 31.0%; Pred. No. 1.8;
Matches 13; Conservative 11; Mismatches 17; Indels 1; Gaps 1;
QY 3 DDDPKRYEDCRRRCMDTRGQKEOQCEESCKSOYGEKDOQ 44
DB 402 NDKAKETE-CNSPKWMDKEKDEKRRKRLSEGRKAERKNE 442
RESULT 5

```

```

MJD1_HUMAN
ID MJD1_HUMAN STANDARD; PRT; 360 AA.
AC P54252;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MACHADO-JOSEPH DISEASE PROTEIN 1.
GN MJD1 OR MJD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=95179166; PubMed=7874163;
RA Kawaguchi Y., Okamoto T., Taniwaki M., Aizawa M., Inoue M.,
RA Katayama S., Kawakami H., Nakamura S., Nishimura M., Akiguchi I.,
RA Kimura J., Natumiya S., Kakizuka A.;
RT "CAG expansions in a novel gene for Machado-Joseph disease at
chromosome 14q32.1."
RL Nat. Genet. 8:221-228(1994).
CC -1- POLYMORPHISM: THE POLY-GLN REGION OF THE MACHADO-JOSEPH PROTEIN IS
CC HIGHLY POLYMORPHIC (14 TO 40 REPEATS) IN THE NORMAL POPULATION AND
CC IS EXPANDED TO ABOUT 68-82 REPEATS IN MJD1 PATIENTS. LONGER
CC EXPANSIONS RESULT IN EARLIER ONSET AND MORE SEVERE CLINICAL
CC MANIFESTATIONS OF THE DISEASE.
CC -1- DISEASE: DEFECTS IN MJD1 ARE THE CAUSE OF MACHADO-JOSEPH DISEASE
CC (MJD), A NEURODEGENERATIVE DISORDER CHARACTERIZED BY CEREBELLAR
CC ATAXIA, PYRAMIDAL AND EXTRAPYRAMIDAL SIGNS, PERIPHERAL NERVE
CC PALSY, EXTERNAL OPHTHALMOPLÉGIA, FACIAL AND LINGUAL FASCICULATION
CC AND BULGING. THIS DISEASE IS AUTOSOMAL AND DOMINANT, WITH A LATE
CC ONSET OF SYMPTOMS, GENERALLY AFTER THE FOURTH DECADE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC
CC EMBL; S75313; AAB33571.1; -.
DR MIM; 109150; -.
DR INTERPRO; IPR002950; -.
DR PFAM; PF02099; Josephin.1.
DR PRINTS; PR01233; JOSEPHIN.
KW Polymorphism; Triplet repeat expansion.
FT DOMAIN 292 317 POLY-GLN.
SQ SEQUENCE 360 AA; 41531 MW; AD6774F29A6CD95D CRC64;

```

Query Match 22.5%; Score 62; DB 1; Length 360;

Best Local Similarity 35.0%; Pred. No. 5.5;

Matches 14; Conservative 12; Mismatches 12; Indels 2; Gaps 1;

OY 10 EDCRRCE--WDTGCKEQQCEESCKSOYGEKQQQRH 47

Db 279 EELRKREAVFEKQKQKQKQKQKQKQKQKQKQK 318

```

RESULT 6
GRPE_HELPY STANDARD; PRT; 189 AA.
ID GRPE_HELPY
AC P55970;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GRPE PROTEIN (HSP-70 COFACTOR).
GN GRPE OR HP0110.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RT Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Maltby L., Mallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori."
RL Nature 388:539-547(1997).
CC -1- FUNCTION: STIMULATES, JOINTLY WITH DNAJ, THE ATPASE ACTIVITY OF
CC DNK. HELPS TO RELEASE ADP FROM DNK THUS ALLOWING DNK TO RECYCLE
CC MORE EFFICIENTLY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GRPE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC
CC EMBL; AE00532; AAD07179.1; -.
DR TIGR; HP0110; -.
DR INTERPRO; IPR000740; -.
DR PFAM; PF01025; GRPE.1.
DR PRINTS; PR00773; GRPEPROTEIN.
DR PROSITE; PS01071; GRPE.1.
KW Chapterone; Heat shock.
SQ SEQUENCE 189 AA; 22040 MW; 62C0991B684F0C2D CRC64;

```

Query Match 22.2%; Score 61; DB 1; Length 189;

Best Local Similarity 44.4%; Pred. No. 3.9;

Matches 12; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 17 EWDTRGCKEQQCEESCKSOYGEKQQ 43

Db 8 EHDHLSQKEPCEKCKEQQYEEKOE 34

```

RESULT 7
TAGB_DICDI STANDARD; PRT; 1905 AA.
ID TAGB_DICDI
AC P54683;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR (EC 3.4.21. -).
GN TAGB.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Dictyostelida; Dictyostelium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=95262903; PubMed=7744252;
RA Shaulsky G., Kuspa A., Loomis W.F.;
RT "A multidrug resistance transporter/serine protease gene is required
for prestalk specialization in Dictyostelium."
RL Genes Dev. 9:1111-1122(1995).
CC -1- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
CC INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
CC S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
CC TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC -1- SIMILARITY: STRONG, TO TAGC.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U20432; AAA62212.1; -.
DR HSSP: P13569; 1NBD.
DR DICTYDB: DD02059; TAGB.
DR INTERPRO: IPR000209; -.
DR INTERPRO: IPR001140; -.
DR INTERPRO: IPR001617; -.
DR PFAM: PF00664; ABC_membrane; 1.
DR PFAM: PF00005; ABC_tran; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
FT SIGNAL 1 ?
FT CHAIN ?
FT DOMAIN 378 700 PRESTALK-SPECIFIC PROTEIN TAGB.
FT DOMAIN 1011 1031 ABC_TRANSPORTER.
FT TRANSMEM 1076 1096 POTENTIAL.
FT TRANSMEM 1121 1141 POTENTIAL.
FT TRANSMEM 1210 1230 POTENTIAL.
FT TRANSMEM 1309 1329 POTENTIAL.
FT TRANSMEM 1332 1352 POTENTIAL.
FT ACT_SITE 387 387 CHARGE_RELAY_SYSTEM (BY SIMILARITY).
FT ACT_SITE 432 432 CHARGE_RELAY_SYSTEM (BY SIMILARITY).
FT ACT_SITE 695 695 CHARGE_RELAY_SYSTEM (BY SIMILARITY).
FT NP_BIND 1553 1560 ATP (POTENTIAL).
FT DOMAIN 63 67 POLY-GLN.
FT DOMAIN 95 104 POLY-ASN.
FT DOMAIN 107 134 POLY-ASN.
FT DOMAIN 311 321 POLY-SER.
FT DOMAIN 833 839 POLY-SER.
FT DOMAIN 838 844 POLY-GLY.
FT DOMAIN 871 876 POLY-LEU.
FT DOMAIN 1012 1015 POLY-ILE.
FT DOMAIN 1386 1389 POLY-GLU.
FT DOMAIN 1398 1404 POLY-GLY.
FT DOMAIN 1445 1450 POLY-ASN.
FT DOMAIN 1765 1779 POLY-ASN.
FT DOMAIN 1782 1785 POLY-SER.
FT DOMAIN 1807 1812 POLY-PRO.
FT DOMAIN 1815 1860 POLY-GLN.
FT DOMAIN 1872 1878 POLY-PRO.
FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1172 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1522 1522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1558 1558 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1905 AA; 212518 MW; B8E223FA8B9AE13C CRC64;

```

Query Match 22.2%; Score 61; DB 1; Length 1905;  
 Best Local Similarity 28.3%; Pred. No. 33;  
 Matches 13; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

OY 2 DDDPKRYEDCRRCEMDTGGKEQOCEESCKSOYGEKDQOQRH 47  
 DB 1805 DDPKRRPPEQOQEQOQOQOQOQOQOQOQOQOQOQOQOQ 1850

```

RESULT 8
GIB1 MAIZE
ID GIB1 MAIZE STANDARD; PRT; 573 AA.
AC P15530;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLOBULIN-1 S ALLELE PRECURSOR (GIB1-S) (7S-LIKE).
GN GIB1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-CV. INBRED LINE VA26;
RA Belanger F.C., Kriz A.L.;
RT "Molecular characterization of the major maize embryo globulin encoded
RT by the GIB1 gene.";
RL Plant Physiol. 91:636-643(1989).
RN 12
RP SEQUENCE OF 87-100.
RX MEDLINE=89374022; PubMed=2775172;
RA Kriz A.L.;
RT "Characterization of embryo globulins encoded by the maize GIB
RT genes.";
RL Biochem. Genet. 27:239-251(1989).
CC -1- PTM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE
CC MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.
CC -1- POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING GIB1 ALLELES HAVE
CC THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL
CC PROTEINS, RESPECTIVELY.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLICANIN, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M24845; AAA33467.1; -.
DR HSSP: P50477; 1CAM.
DR MAIZEDB: 30181; -.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 18 OR 21 (POTENTIAL).
FT PROPEP 19 86
FT CHAIN 87 573 GLOBULIN-1 S ALLELE.
FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 573 AA; 65029 MW; 525ED1D0A062976 CMC64;

```

Query Match 21.6%; Score 59.5; DB 1; Length 573;  
 Best Local Similarity 38.9%; Pred. No. 16;  
 Matches 14; Conservative 7; Mismatches 12; Indels 3; Gaps 2;

OY 12 CRRCEMDTGGKEQOCEESCKSOYGEKDQOQRH 46  
 DB 40 CVRCE--DRPHQRPCLCEQCEEREKERKERSRH 73

RESULT 9  
 DPOA\_TRYBB  
 ID DPOA\_TRYBB STANDARD; PRT; 1339 AA.  
 AC P27727;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA POLYMERASE ALPHA CATALYTIC SUBUNIT (EC 2.7.7.7) (DNA POLYMERASE

```

DE 1).
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5427.
RX MEDLINE=92093600; PubMed=1754381;
RA Leegwater P.A.J., Strating M.S., Murphy N.B., Kooy R.F.,
  van der Vliet P.C., Overduive J.P.;
RT "The Trypanosoma brucei DNA polymerase alpha core subunit gene is
  developmentally regulated and linked to a constitutively expressed
  open reading frame".
RL Nucleic Acids Res. 19:6441-6447(1991).
CC -1- FUNCTION: POLYMERASE ALPHA IN A COMPLEX WITH DNA PRIMASE IS A
  REPLICATIVE POLYMERASE.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
  N PYROPHOSPHATE + DNA(N).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
  ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
  DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-slb.ch/announce/
  or send an email to license@isb-slb.ch).
CC -----
CC EMBL: X60951; CAA43287.1; -.
CC PIR: S20052; S20052.
CC INTERPRO: IPR002064; -.
CC PFAM: PF00136; DNA_POL_B.1.
CC PRINTS: PR00106; DNAPOLB.
CC PROSITE: PS00116; DNA_POLYMERASE_B.1.
CC TRANSFERASE; DNA-directed DNA polymerase; DNA replication;
  DNA-binding; Nuclear protein.
CC KW SEQUENCE 1339 AA; 151611 MW; 9FF159412F2B7FBA CRC64;
SO

```

Query Match 21.5%; Score 59; DB 1; Length 1339;  
 Best Local Similarity 32.3%; Pred. No. 39;  
 Matches 10; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

```

QY 17 EMDTRGQKGGQCEESCKSGYGGKDDQQRHR 47
DB 3 DWVSCRSEGRKCEKGGSTFDESEEEQWR 33

```

RESULT 10  
 FBL4\_CRIGR STANDARD: PRT; 443 AA.  
 AC 055058;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR MATRIX PROTEIN 2 PRECURSOR  
 DE (FIBULIN-4) (FIBL-4) (HALL PROTEIN).  
 GN EFEMP2 OR FBLN4.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OVARY;  
 RA Heine H., Delude R.L., Monks B., Golenbock D.T.;  
 RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.  
 CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-slb.ch/announce/
  or send an email to license@isb-slb.ch).
CC -----
CC EMBL: AF046870; AAC03101.1; -.
CC HSSP: P00736; IAPQ.
CC INTERPRO: IPR000152; -.
CC INTERPRO: IPR000561; -.
CC INTERPRO: IPR001881; -.
CC PFAM: PF00008; EGF_4.
CC PROSITE: PS00010; ASX_HYDROXYL_4.
CC PROSITE: PS00022; EGF_1; FALSE_NEG.
CC PROSITE: PS01186; EGF_2; 4.
CC PROSITE: PS01187; EGF_CA; 6.
CC Repeat; EGF-like domain; Calcium-binding; glycoprotein; signal.
CC KW Repeat; EGF-like domain; Calcium-binding; glycoprotein; signal.
CC FT CHAIN 1 25
  FT 1 26 443
  FT 123 163 81
  FT 164 202 162
  FT 203 242 177
  FT 243 282 151
  FT 283 328 162
  FT 329 367 149
  FT 368 406 162
  FT 407 443 177
  FT 444 482 173
  FT 483 521 186
  FT 522 560 201
  FT 561 599 217
  FT 600 638 226
  FT 639 677 241
  FT 678 716 258
  FT 717 755 267
  FT 756 794 281
  FT 795 833 300
  FT 834 872 287
  FT 873 911 294
  FT 912 950 309
  FT 951 989 315
  FT 990 1028 327
  FT 1029 1067 394
  FT 1068 1106 394
  FT 1107 1145 394
  FT 1146 1184 394
  FT 1185 1223 394
  FT 1224 1262 394
  FT 1263 1301 394
  FT 1302 1340 394
  FT 1341 1379 394
  FT 1380 1418 394
  FT 1419 1457 394
  FT 1458 1496 394
  FT 1497 1535 394
  FT 1536 1574 394
  FT 1575 1613 394
  FT 1614 1652 394
  FT 1653 1691 394
  FT 1692 1730 394
  FT 1731 1769 394
  FT 1770 1808 394
  FT 1809 1847 394
  FT 1848 1886 394
  FT 1887 1925 394
  FT 1926 1964 394
  FT 1965 2003 394
  FT 2004 2042 394
  FT 2043 2081 394
  FT 2082 2120 394
  FT 2121 2159 394
  FT 2160 2198 394
  FT 2199 2237 394
  FT 2238 2276 394
  FT 2277 2315 394
  FT 2316 2354 394
  FT 2355 2393 394
  FT 2394 2432 394
  FT 2433 2471 394
  FT 2472 2510 394
  FT 2511 2549 394
  FT 2550 2588 394
  FT 2589 2627 394
  FT 2628 2666 394
  FT 2667 2705 394
  FT 2706 2744 394
  FT 2745 2783 394
  FT 2784 2822 394
  FT 2823 2861 394
  FT 2862 2900 394
  FT 2901 2939 394
  FT 2940 2978 394
  FT 2979 3017 394
  FT 3018 3056 394
  FT 3057 3095 394
  FT 3096 3134 394
  FT 3135 3173 394
  FT 3174 3212 394
  FT 3213 3251 394
  FT 3252 3290 394
  FT 3291 3329 394
  FT 3330 3368 394
  FT 3369 3407 394
  FT 3408 3446 394
  FT 3447 3485 394
  FT 3486 3524 394
  FT 3525 3563 394
  FT 3564 3602 394
  FT 3603 3641 394
  FT 3642 3680 394
  FT 3681 3719 394
  FT 3720 3758 394
  FT 3759 3797 394
  FT 3798 3836 394
  FT 3837 3875 394
  FT 3876 3914 394
  FT 3915 3953 394
  FT 3954 3992 394
  FT 3993 4031 394
  FT 4032 4070 394
  FT 4071 4109 394
  FT 4110 4148 394
  FT 4149 4187 394
  FT 4188 4226 394
  FT 4227 4265 394
  FT 4266 4304 394
  FT 4305 4343 394
  FT 4344 4382 394
  FT 4383 4421 394
  FT 4422 4460 394
  FT 4461 4499 394
  FT 4500 4538 394
  FT 4539 4577 394
  FT 4578 4616 394
  FT 4617 4655 394
  FT 4656 4694 394
  FT 4695 4733 394
  FT 4734 4772 394
  FT 4773 4811 394
  FT 4812 4850 394
  FT 4851 4889 394
  FT 4890 4928 394
  FT 4929 4967 394
  FT 4968 5005 394
  FT 5006 5044 394
  FT 5045 5083 394
  FT 5084 5122 394
  FT 5123 5161 394
  FT 5162 5200 394
  FT 5201 5239 394
  FT 5240 5278 394
  FT 5279 5317 394
  FT 5318 5356 394
  FT 5357 5395 394
  FT 5396 5434 394
  FT 5435 5473 394
  FT 5474 5512 394
  FT 5513 5551 394
  FT 5552 5590 394
  FT 5591 5629 394
  FT 5630 5668 394
  FT 5669 5707 394
  FT 5708 5746 394
  FT 5747 5785 394
  FT 5786 5824 394
  FT 5825 5863 394
  FT 5864 5902 394
  FT 5903 5941 394
  FT 5942 5980 394
  FT 5981 6019 394
  FT 6020 6058 394
  FT 6059 6097 394
  FT 6098 6135 394
  FT 6136 6174 394
  FT 6175 6213 394
  FT 6214 6252 394
  FT 6253 6291 394
  FT 6292 6330 394
  FT 6331 6368 394
  FT 6369 6407 394
  FT 6408 6446 394
  FT 6447 6485 394
  FT 6486 6523 394
  FT 6524 6562 394
  FT 6563 6601 394
  FT 6602 6640 394
  FT 6641 6678 394
  FT 6679 6717 394
  FT 6718 6756 394
  FT 6757 6795 394
  FT 6796 6834 394
  FT 6835 6873 394
  FT 6874 6912 394
  FT 6913 6951 394
  FT 6952 6989 394
  FT 6990 7028 394
  FT 7029 7067 394
  FT 7068 7106 394
  FT 7107 7145 394
  FT 7146 7184 394
  FT 7185 7223 394
  FT 7224 7262 394
  FT 7263 7301 394
  FT 7302 7340 394
  FT 7341 7379 394
  FT 7380 7418 394
  FT 7419 7457 394
  FT 7458 7496 394
  FT 7497 7535 394
  FT 7536 7574 394
  FT 7575 7613 394
  FT 7614 7652 394
  FT 7653 7691 394
  FT 7692 7730 394
  FT 7731 7769 394
  FT 7770 7808 394
  FT 7809 7847 394
  FT 7848 7886 394
  FT 7887 7925 394
  FT 7926 7964 394
  FT 7965 8003 394
  FT 8004 8042 394
  FT 8043 8081 394
  FT 8082 8120 394
  FT 8121 8159 394
  FT 8160 8198 394
  FT 8199 8237 394
  FT 8238 8276 394
  FT 8277 8315 394
  FT 8316 8354 394
  FT 8355 8393 394
  FT 8394 8432 394
  FT 8433 8471 394
  FT 8472 8510 394
  FT 8511 8549 394
  FT 8550 8588 394
  FT 8589 8627 394
  FT 8628 8666 394
  FT 8667 8705 394
  FT 8706 8744 394
  FT 8745 8783 394
  FT 8784 8822 394
  FT 8823 8861 394
  FT 8862 8900 394
  FT 8901 8939 394
  FT 8940 8978 394
  FT 8979 9017 394
  FT 9018 9056 394
  FT 9057 9095 394
  FT 9096 9134 394
  FT 9135 9173 394
  FT 9174 9212 394
  FT 9213 9251 394
  FT 9252 9290 394
  FT 9291 9329 394
  FT 9330 9368 394
  FT 9369 9407 394
  FT 9408 9446 394
  FT 9447 9485 394
  FT 9486 9524 394
  FT 9525 9563 394
  FT 9564 9602 394
  FT 9603 9641 394
  FT 9642 9680 394
  FT 9681 9719 394
  FT 9720 9758 394
  FT 9759 9797 394
  FT 9798 9836 394
  FT 9837 9875 394
  FT 9876 9914 394
  FT 9915 9953 394
  FT 9954 9992 394
  FT 10000 10000 394
  FT 10001 10000 394
  FT 10002 10000 394
  FT 10003 10000 394
  FT 10004 10000 394
  FT 10005 10000 394
  FT 10006 10000 394
  FT 10007 10000 394
  FT 10008 10000 394
  FT 10009 10000 394
  FT 10010 10000 394
  FT 10011 10000 394
  FT 10012 10000 394
  FT 10013 10000 394
  FT 10014 10000 394
  FT 10015 10000 394
  FT 10016 10000 394
  FT 10017 10000 394
  FT 10018 10000 394
  FT 10019 10000 394
  FT 10020 10000 394
  FT 10021 10000 394
  FT 10022 10000 394
  FT 10023 10000 394
  FT 10024 10000 394
  FT 10025 10000 394
  FT 10026 10000 394
  FT 10027 10000 394
  FT 10028 10000 394
  FT 10029 10000 394
  FT 10030 10000 394
  FT 10031 10000 394
  FT 10032 10000 394
  FT 10033 10000 394
  FT 10034 10000 394
  FT 10035 10000 394
  FT 10036 10000 394
  FT 10037 10000 394
  FT 10038 10000 394
  FT 10039 10000 394
  FT 10040 10000 394
  FT 10041 10000 394
  FT 10042 10000 394
  FT 10043 10000 394
  FT 10044 10000 394
  FT 10045 10000 394
  FT 10046 10000 394
  FT 10047 10000 394
  FT 10048 10000 394
  FT 10049 10000 394
  FT 10050 10000 394
  FT 10051 10000 394
  FT 10052 10000 394
  FT 10053 10000 394
  FT 10054 10000 394
  FT 10055 10000 394
  FT 10056 10000 394
  FT 10057 10000 394
  FT 10058 10000 394
  FT 10059 10000 394
  FT 10060 10000 394
  FT 10061 10000 394
  FT 10062 10000 394
  FT 10063 10000 394
  FT 10064 10000 394
  FT 10065 10000 394
  FT 10066 10000 394
  FT 10067 10000 394
  FT 10068 10000 394
  FT 10069 10000 394
  FT 10070 10000 394
  FT 10071 10000 394
  FT 10072 10000 394
  FT 10073 10000 394
  FT 10074 10000 394
  FT 10075 10000 394
  FT 10076 10000 394
  FT 10077 10000 394
  FT 10078 10000 394
  FT 10079 10000 394
  FT 10080 10000 394
  FT 10081 10000 394
  FT 10082 10000 394
  FT 10083 10000 394
  FT 10084 10000 394
  FT 10085 10000 394
  FT 10086 10000 394
  FT 10087 10000 394
  FT 10088 10000 394
  FT 10089 10000 394
  FT 10090 10000 394
  FT 10091 10000 394
  FT 10092 10000 394
  FT 10093 10000 394
  FT 10094 10000 394
  FT 10095 10000 394
  FT 10096 10000 394
  FT 10097 10000 394
  FT 10098 10000 394
  FT 10099 10000 394
  FT 10100 10000 394
  FT 10101 10000 394
  FT 10102 10000 394
  FT 10103 10000 394
  FT 10104 10000 394
  FT 10105 10000 394
  FT 10106 10000 394
  FT 10107 10000 394
  FT 10108 10000 394
  FT 10109 10000 394
  FT 10110 10000 394
  FT 10111 10000 394
  FT 10112 10000 394
  FT 10113 10000 394
  FT 10114 10000 394
  FT 10115 10000 394
  FT 10116 10000 394
  FT 10117 10000 394
  FT 10118 10000 394
  FT 10119 10000 394
  FT 10120 10000 394
  FT 10121 10000 394
  FT 10122 10000 394
  FT 10123 10000 394
  FT 10124 10000 394
  FT 10125 10000 394
  FT 10126 10000 394
  FT 10127 10000 394
  FT 10128 10000 394
  FT 10129 10000 394
  FT 10130 10000 394
  FT 10131 10000 394
  FT 10132 10000 394
  FT 10133 10000 394
  FT 10134 10000 394
  FT 10135 10000 394
  FT 10136 10000 394
  FT 10137 10000 394
  FT 10138 10000 394
  FT 10139 10000 394
  FT 10140 10000 394
  FT 10141 10000 394
  FT 10142 10000 394
  FT 10143 10000 394
  FT 10144 10000 394
  FT 10145 10000 394
  FT 10146 10000 394
  FT 10147 10000 394
  FT 10148 10000 394
  FT 10149 10000 394
  FT 10150 10000 394
  FT 10151 10000 394
  FT 10152 10000 394
  FT 10153 10000 394
  FT 10154 10000 394
  FT 10155 10000 394
  FT 10156 10000 394
  FT 10157 10000 394
  FT 10158 10000 394
  FT 10159 10000 394
  FT 10160 10000 394
  FT 10161 10000 394
  FT 10162 10000 394
  FT 10163 10000 394
  FT 10164 10000 394
  FT 10165 10000 394
  FT 10166 10000 394
  FT 10167 10000 394
  FT 10168 10000 394
  FT 10169 10000 394
  FT 10170 10000 394
  FT 10171 10000 394
  FT 10172 10000 394
  FT 10173 10000 394
  FT 10174 10000 394
  FT 10175 10000 394
  FT 10176 10000 394
  FT 10177 10000 394
  FT 10178 10000 394
  FT 10179 10000 394
  FT 10180 10000 394
  FT 10181 10000 394
  FT 10182 10000 394
  FT 10183 10000 394
  FT 10184 10000 394
  FT 10185 10000 394
  FT 10186 10000 394
  FT 10187 10000 394
  FT 10188 10000 394
  FT 10189 10000 394
  FT 10190 10000 394
  FT 10191 10000 394
  FT 10192 10000 394
  FT 10193 10000 394
  FT 10194 10000 394
  FT 10195 10000 394
  FT 10196 10000 394
  FT 10197 10000 394
  FT 10198 10000 394
  FT 10199 10000 394
  FT 10200 10000 394
  FT 10201 10000 394
  FT 10202 10000 394
  FT 10203 10000 394
  FT 10204 10000 394
  FT 10205 10000 394
  FT 10206 10000 394
  FT 10207 10000 394
  FT 10208 10000 394
  FT 10209 10000 394
  FT 10210 10000 394
  FT 10211 10000 394
  FT 10212 10000 394
  FT 10213 10000 394
  FT 10214 10000 394
  FT 10215 10000 394
  FT 10216 10000 394
  FT 10217 10000 394
  FT 10218 10000 394
  FT 10219 10000 394
  FT 10220 10000 394
  FT 10221 10000 394
  FT 10222 10000 394
  FT 10223 10000 394
  FT 10224 10000 394
  FT 10225 10000 394
  FT 10226 10000 394
  FT 10227 10000 394
  FT 10228 10000 394
  FT 10229 10000 394
  FT 10230 10000 394
  FT 10231 10000 394
  FT 10232 10000 394
  FT 10233 10000 394
  FT 10234 10000 394
  FT 10235 10000 394
  FT 10236 10000 394
  FT 10237 10000 394
  FT 10238 10000 394
  FT 10239 10000 394
  FT 10240 10000 394
  FT 10241 10000 394
  FT 10242 10000 394
  FT 10243 10000 394
  FT 10244 10000 394
  FT 10245 10000 394
  FT 10246 10000 394
  FT 10247 10000 394
  FT 10248 10000 394
  FT 10249 10000 394
  FT 10250 10000 394
  FT 10251 10000 394
  FT 10252 10000 394
  FT 10253 10000 394
  FT 10254 10000 394
  FT 10255 10000 394
  FT 10256 10000 394
  FT 10257 10000 394
  FT 10258 10000 394
  FT 10259 10000 394
  FT 10260 10000 394
  FT 10261 10000 394
  FT 10262 10000 394
  FT 10263 10000 394
  FT 10264 10000 394
  FT 10265 10000 394
  FT 10266 10000 394
  FT 10267 10000 394
  FT 10268 10000 394
  FT 10269 10000 394
  FT 10270 10000 394
  FT 10271 10000 394
  FT 10272 10000 394
  FT 10273 10000 394
  FT 10274 10000 394
  FT 10275 10000 394
  FT 10276 10000 394
  FT 10277 10000 394
  FT 10278 10000 394
  FT 10279 10000 394
  FT 10280 10000 394
  FT 10281 10000 394
  FT 10282 10000 394
  FT 10283 10000 394
  FT 10284 10000 394
  FT 10285 10000 394
  FT 10286 10000 394
  FT 10287 10000 394
  FT 10288 10000 394
  FT 10289 10000 394
  FT 10290 10000 394
  FT 10291 10000 394
  FT 10292 10000 394
  FT 10293 10000 394
  FT 10294 10000 394
  FT 10295 10000 394
  FT 10296 10000 394
  FT 10297 10000 394
  FT 10298 10000 394
  FT 10299 10000 394
  FT 10300 10000 394
  FT 10301 10000 394
  FT 10302 10000 394
  FT 10303 10000 394
  FT 10304 10000 394
  FT 10305 10000 394
  FT 10306 10000 394
  FT 10307 10000 394
  FT 10308 10000 394
  FT 10309 10000 394
  FT 10310 10000 394
  FT 10311 10000 394
  FT 10312 10000 394
  FT 10313 10000 394
  FT 10314 10000 394
  FT 10315 10000 394
  FT 10316 10000 394
  FT 10317 10000 394
  FT 10318 10000 394
  FT 10319 10000 394
  FT 10320 10000 394
  FT 10321 10000 394
  FT 10322 10000 394
  FT 10323 10000 394
  FT 10324 10000 394
  FT 10325 10000 394
  FT 10326 10000 394
  FT 10327 10000 394
  FT 10328 10000 394
  FT 10329 10000 394
  FT
```





```
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 443 EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR
FT DOMAIN 36 81 MATRIX PROTEIN 2.
FT DOMAIN 123 163 EGF-LIKE 1, DIVERGENT.
FT DOMAIN 164 202 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 203 242 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 243 282 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 283 328 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 328 340 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DISULFID 127 140 BY SIMILARITY.
FT DISULFID 134 149 BY SIMILARITY.
FT DISULFID 151 162 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 188 201 BY SIMILARITY.
FT DISULFID 207 217 BY SIMILARITY.
FT DISULFID 213 226 BY SIMILARITY.
FT DISULFID 228 241 BY SIMILARITY.
FT DISULFID 247 258 BY SIMILARITY.
FT DISULFID 254 267 BY SIMILARITY.
FT DISULFID 269 281 BY SIMILARITY.
FT DISULFID 287 300 BY SIMILARITY.
FT DISULFID 294 309 BY SIMILARITY.
FT DISULFID 315 327 BY SIMILARITY.
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 443 AA: 49425 MM: 4969C0328A23DD88 CRC64:
```

```
Query Match 20.9%; Score 57.5; DB 1; Length 443;
Best Local Similarity 30.0%; Pred. No. 20;
Matches 12; Conservative 7; Mismatches 16; Indels 5; Gaps 2;
```

```
OY 2 DDDPPRYEDCRRRCFMDTRGQ--KEOOC---ESCKSQ 36
| : | | : | | | : | : | : | : | : | : |
DB 29 DPEPDSYTECTDGYEMDADSDCHCRDVNCELTIPACKGE 68
```

Search completed: March 1, 2001, 16:16:55  
Job time: 435 sec